# A representation theorem for (trAp)1/p

Magnus, J.R.

*Published in:*
Linear Algebra and its Applications

*Publication date:*
1987

# A Representation Theorem for $(\operatorname{tr} A^p)^{1/p}$

Jan R. Magnus
*London School of Economics*
*Houghton Street*
*London, WC2A 2AE England*

Submitted by Ingram Olkin

ABSTRACT

An inequality for positive semidefinite matrices is proved, and from it a quasilinear representation of $(\operatorname{tr} A^p)^{1/p}$ is obtained. From this representation follow matrix versions of the inequalities of Hölder and Minkowski.

## 1. INTRODUCTION

The fundamental inequalities of Hölder and Minkowski can be proved in a variety of ways; see Hardy, Littlewood, and Pólya (1952). A particularly interesting proof of both inequalities is based on the following lemma, of interest in itself, which is easily established by the method of Lagrange or otherwise.

LEMMA 1.[1] *Let $p > 1$, $q = p/(p-1)$, and $a_i \geqslant 0$ $(i = 1, \dots, n)$. Then*

$$\sum_{i=1}^{n} a_i x_i \leqslant \left( \sum_{i=1}^{n} a_i^p \right)^{1/p} \tag{1.1}$$

*for every set of nonnegative quantities $x_1, x_2, \dots, x_n$ satisfying $\sum_{i=1}^{n} x_i^q = 1$. Equality in (1.1) occurs if and only if $a_1 = a_2 = \cdots = a_n = 0$ or $x_i^q = a_i^p / \sum_{j=1}^{n} a_j^p$ $(i = 1, \dots, n)$.*

[1]See Beckenbach and Bellman (1961, Theorem 5, p. 23).




52 Vanderbilt Ave., New York, NY 10017 0024-3795/87/$3.50

An immediate and easy consequence of Lemma 1 is

HÖLDER'S INEQUALITY. *If $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$ are vectors in the nonnegative orthant of $\mathbb{R}^n$ and $0<\alpha<1$, then*

$$\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha} \leqslant \left(\sum_{i=1}^{n} x_i\right)^{\alpha}\left(\sum_{i=1}^{n} y_i\right)^{1-\alpha} \tag{1.2}$$

*with equality if and only if $x$ and $y$ are linearly dependent.*

If we formalate the inequality (1.1) as a maximization problem, we get

$$\max_R \sum_{i=1}^{n} a_i x_i = \left(\sum_{i=1}^{n} a_i^p\right)^{1/p}, \tag{1.3}$$

where $R$ is the region defined by

$$\sum_{i=1}^{n} x_i^q = 1, \qquad x_i \geqslant 0.$$

The idea of expressing a *nonlinear* function, such as $(\Sigma a_i^p)^{1/p}$, as an envelope of *linear* functions goes back to Minkowski and was used extensively by Bellman and others. This technique is called *quasilinearization*. A direct consequence of the quasilinear representation (1.3) is

MINKOWSKI'S INEQUALITY. *If $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$ are vectors in the nonnegative orthant of $\mathbb{R}^n$ and $p>1$, then*

$$\left(\sum_{i=1}^{n} (x_i+y_i)^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p} \tag{1.4}$$

*with equality if and only if $x$ and $y$ are linearly dependent.*

The purpose of this paper is to extend the inequality (1.1) to positive semidefinite matrices, thus obtaining a quasilinear representation of $(\operatorname{tr} A^p)^{1/p}$. This is achieved in Theorem 5. Interesting matrix versions of the inequalities of Hölder and Minkowski are then easily derived from this representation (Theorems 6 and 7). Several preliminary results, of interest in themselves, are also reported.

## 2. A NECESSARY AND SUFFICIENT CONDITION FOR DIAGONALITY

Let us begin by stating the following simple but useful result concerning the conditions under which a real symmetric (or positive definite) matrix is diagonal.

THEOREM 1. *A positive definite matrix is diagonal if and only if the product of its diagonal elements is equal to its determinant. A real symmetric matrix is diagonal if and only if its diagonal elements and its eigenvalues coincide.*

*Proof.* A proof by induction and perturbation is straightforward and therefore omitted. A variety of other proofs can also be constructed. ■

## 3. KARAMATA'S INEQUALITY

Let $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ be two vectors in $\mathbb{R}^n$. We shall say that $x$ is *majorized* by $y$, and write

$$(x_1, \dots, x_n) \prec (y_1, \dots, y_n),$$

when the following three relations are satisfied:

$$x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n;$$
$$x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n, \qquad y_1 \geqslant y_2 \geqslant \cdots \geqslant y_n;$$
$$x_1 + x_2 + \cdots + x_k \leqslant y_1 + y_2 + \cdots + y_k \qquad (1 \leqslant k \leqslant n-1).$$

The theory of majorization originated with the work of Schur (1923) and Hardy, Littlewood, and Pólya (1929), and is now firmly established; see Marshall and Olkin (1979). The following theorem was essentially proved by Schur (1923). A continuous analogue was proved by Hardy, Littlewood, and Pólya (1929), and an important generalization provided by Karamata (1932).

THEOREM 2. *Let $\phi$ be a real-valued convex function defined on an interval $I \subset \mathbb{R}$. If $(x_1, \dots, x_n) \prec (y_1, \dots, y_n)$ on $I^n$, then*

$$\sum_{i=1}^{n} \phi(x_i) \leqslant \sum_{i=1}^{n} \phi(y_i). \tag{3.1}$$

*If, in addition, $\phi$ is strictly convex on I, then equality in* (3.1) *occurs if and only if* $x_i = y_i$ $(i = 1, \ldots, n)$.

*Proof.* See Marshall and Olkin (1979, Propositions 3.C.1 and 3.C.1.a.). ■

An important application of Karamata's inequality is given in Theorem 3.

THEOREM 3. *Let* $A = (a_{ij})$ *be a real symmetric* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *Then for any convex function* $\phi: \mathbb{R} \to \mathbb{R}$ *we have*

$$\sum_{i=1}^{n} \phi(a_{ii}) \leqslant \sum_{i=1}^{n} \phi(\lambda_i). \tag{3.2}$$

*Moreover, if $\phi$ is strictly convex, then equality in* (3.2) *occurs if and only if A is diagonal.*

*Proof.* Without loss of generality we may assume that

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n, \qquad a_{11} \geqslant a_{22} \geqslant \cdots \geqslant a_{nn}.$$

It is well known that $(a_{11}, \ldots, a_{nn})$ is majorized by $(\lambda_1, \ldots, \lambda_n)$; see Schur (1923) or Marshall and Olkin (1979, Theorem 9.B.1). Thus (3.2) follows from Theorem 2. For strictly convex $\phi$, Theorem 2 implies that equality in (3.2) holds if and only if $a_{ii} = \lambda_i$ for $i = 1, \ldots, n$, and, by Theorem 1, this is the case if and only if $A$ is diagonal. ■

As a corollary of Theorem 3, let us establish the following useful inequalities concerning positive semidefinite matrices, which we shall use in proving Theorem 5.

THEOREM 4. *Let* $A = (a_{ij})$ *be a positive semidefinite* $n \times n$ *matrix. Then*

$$\operatorname{tr} A^p \geqslant \sum_{i=1}^{n} a_{ii}^p \qquad (p > 1)$$

*and*

$$\operatorname{tr} A^p \leqslant \sum_{i=1}^{n} a_{ii}^p \qquad (0 < p < 1)$$

*with equality if and only if A is diagonal.*

*Proof.* Let $p > 1$, and define $\phi(x) = x^p$ $(x \geqslant 0)$. The function $\phi$ is strictly convex. Hence Theorem 3 implies that

$$\operatorname{tr} A^p = \sum_{i=1}^{n} \lambda_i^p(A) = \sum_{i=1}^{n} \phi(\lambda_i(A)) \geqslant \sum_{i=1}^{n} \phi(a_{ii}) = \sum_{i=1}^{n} a_{ii}^p$$

with equality if and only if $A$ is diagonal. Next, let $0 < p < 1$. Then $\psi(x) \equiv -x^p$ $(x \geqslant 0)$ is strictly convex, and the second result follows in the same way. ■

## 4. QUASILINEAR REPRESENTATION OF $(\operatorname{tr} A^p)^{1/p}$

We now have all the ingredients to prove the matrix analogue of Lemma 1.

THEOREM 5. *Let $p > 1$, $q = p/(p-1)$, and let $A \neq 0$ be a positive semidefinite $n \times n$ matrix. Then*

$$\operatorname{tr} AX \leqslant (\operatorname{tr} A^p)^{1/p} \tag{4.1}$$

*for every positive semidefinite $n \times n$ matrix $X$ satisfying $\operatorname{tr} X^q = 1$. Equality in (4.1) occurs if and only if*

$$X^q = \frac{1}{\operatorname{tr} A^p} A^p. \tag{4.2}$$

REMARK. If we let $R$ be the region

$$R = \{X : X \in \mathbb{R}^{n \times n},\ X \text{ positive semidefinite}, \operatorname{tr} X^q = 1\},$$

then we can state Theorem 5 equivalently as

$$\max_R \operatorname{tr} AX = (\operatorname{tr} A^p)^{1/p} \tag{4.3}$$

for every positive semidefinite $n \times n$ matrix $A$. Thus we can express $(\operatorname{tr} A^p)^{1/p}$ (a *nonlinear* function of $A$) as an envelope of *linear* functions of $A$. In other words, we obtain a quasilinear representation of $(\operatorname{tr} A^p)^{1/p}$.

*Proof.* Let $X$ be an arbitrary positive semidefinite $n \times n$ matrix satisfying $\operatorname{tr} X^q = 1$. Let $S$ be an orthogonal matrix such that $S'XS = \Lambda$, where $\Lambda$ is diagonal and has the eigenvalues of $X$ as its diagonal elements. Define $B = (b_{ij}) = S'AS$. Then

$$\operatorname{tr} AX = \operatorname{tr} B\Lambda = \sum b_{ii}\lambda_i$$

and

$$\operatorname{tr} X^q = \operatorname{tr} \Lambda^q = \sum \lambda_i^q.$$

Hence, by Lemma 1,

$$\operatorname{tr} AX = \sum b_{ii}\lambda_i \leqslant \left(\sum b_{ii}^p\right)^{1/p}. \tag{4.4}$$

Since $A$ is positive semidefinite, so is $B$, and Theorem 4 thus implies that

$$\sum b_{ii}^p \leqslant \operatorname{tr} B^p. \tag{4.5}$$

Combining (4.4) and (4.5) we obtain

$$\operatorname{tr} AX \leqslant (\operatorname{tr} B^p)^{1/p} = (\operatorname{tr} A^p)^{1/p}.$$

Equality in (4.4) occurs if and only if

$$\lambda_i^q = \frac{b_{ii}^p}{\Sigma b_{ii}^p} \qquad (i = 1, \ldots, n),$$

and equality in (4.5) if and only if $B$ is diagonal. Hence, equality in (4.1) occurs if and only if

$$\Lambda^q = \frac{B^p}{\operatorname{tr} B^p},$$

which is equivalent to (4.2). This concludes the proof. ■

## 5. MATRIX ANALOGUES OF THE INEQUALITIES OF HÖLDER AND MINKOWSKI

An immediate consequence of Theorem 5 is the matrix analogue of Hölder's inequality (1.2).

THEOREM 6. *For any two positive semidefinite matrices $A$ and $B$ of the same order, $A \neq 0$, $B \neq 0$, and $0 < \alpha < 1$, we have*

$$\operatorname{tr} A^{\alpha} B^{1-\alpha} \leqslant (\operatorname{tr} A)^{\alpha} (\operatorname{tr} B)^{1-\alpha}, \tag{5.1}$$

*with equality if and only if $B = \mu A$ for some scalar $\mu > 0$.*

*Proof.* Let $p = 1/\alpha$, $q = 1/(1-\alpha)$, and assume $B \neq 0$. Now define

$$X = \frac{B^{1/q}}{(\operatorname{tr} B)^{1/q}}.$$

Then $\operatorname{tr} X^q = 1$, and hence Theorem 5 applied to $A^{1/p}$ yields

$$\operatorname{tr} A^{1/p} B^{1/q} \leqslant (\operatorname{tr} A)^{1/p} (\operatorname{tr} B)^{1/q},$$

which is (5.1). According to Theorem 5, equality in (5.1) can only occur if $X^q = (1/\operatorname{tr} A) A$, that is, if $B = \mu A$ for some $\mu > 0$. ∎

Another consequence of Theorem 5, more specifically of its quasilinear representation (4.3), is the matrix version of Minkowski's inequality (1.4).

THEOREM 7. *For any two positive semidefinite matrices $A$ and $B$ of the same order ($A \neq 0$, $B \neq 0$), and $p > 1$, we have*

$$\left[\operatorname{tr}(A+B)^p\right]^{1/p} \leqslant (\operatorname{tr} A^p)^{1/p} + (\operatorname{tr} B^p)^{1/p}. \tag{5.2}$$

*with equality if and only if $A = \mu B$ for some $\mu > 0$.*

*Proof.* Let $p > 1$, $q = p/(p-1)$, and let $R$ be the region

$$R = \left\{ X : X \in \mathbb{R}^{n \times n},\ X \text{ positive semidefinite, } \operatorname{tr} X^q = 1 \right\}.$$

Using the quasilinear representation (4.3) of $(\operatorname{tr} A^p)^{1/p}$,

$$\max_R \operatorname{tr} AX = \left(\operatorname{tr} A^p\right)^{1/p},$$

we obtain

$$\begin{aligned} \left[\operatorname{tr}(A+B)^p\right]^{1/p} &= \max_R \operatorname{tr}(A+B)X \\ &\leqslant \max_R \operatorname{tr} AX + \max_R \operatorname{tr} BX \\ &= \left(\operatorname{tr} A^p\right)^{1/p} + \left(\operatorname{tr} B^p\right)^{1/p}. \end{aligned} \tag{5.3}$$

Equality in (5.3) can only occur if the same $X$ maximizes $\operatorname{tr} AX$, $\operatorname{tr} BX$, and $\operatorname{tr}(A+B)X$, which implies, by Theorem 5, that $A^p$, $B^p$, and $(A+B)^p$ are proportional, and hence that $A$ and $B$ are proportional. ■

*I am very grateful to Ken Binmore for his constructive remarks.*

## REFERENCES


Beckenbach, E. F. and Bellman, R. 1961. *Inequalities*, Springer, Berlin.

Hardy, G. H., Littlewood, J. E., and Pólya, G. 1929. Some simple inequalities satisfied by convex functions, *Messenger Math.* 58:145–152.

Hardy, G. H., Littlewood, J. E., and Pólya, G. 1952. *Inequalities* (2nd ed.), Cambridge U.P., Cambridge.

Karamata, J. 1932. Sur une inégalité rélative aux fonctions convexes, *Publ. Math. Univ. Belgrade* 1:145–148.

Marshall, A. W. and Olkin, I. 1979. *Inequalities: Theory of Majorization and Its Applications*, Academic, New York.

Schur, I. 1923. Ueber eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie, *Sitzungsber. Berlin. Math. Ges.* 22:9–20.


*Received 23 December 1985; revised 17 November 1986*